

Mathematics Is Biology's Next Microscope, Only Better; Biology Is Mathematics' Next Physics, Only Better

Joel E. Cohen
Laboratory of Populations
Rockefeller & Columbia Universities
12 February 2003

Text Copyright © 2003 by Joel E. COHEN

All images are copyrighted by their owners, who have granted permission for their use, or are public domain. Any use of copyrighted material must be negotiated between the user and the copyright owner if any.

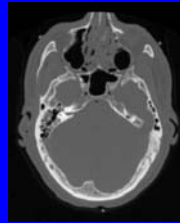
This is a revised version of slides presented Feb. 12, 2003, as the keynote address of the Symposium on Accelerating Mathematical-Biological Linkages, Natcher Building, National Institutes of Health.

The text of the slides has been substantially improved by the comments of many members of the Symposium (see acknowledgments at end). The notes that accompany each slide may contain materials quoted from others; in such cases, authors are identified.

Thanks to Margaret Palmer, Sam Scheiner, Michael Steuerwalt, James Cassatt, Mike Marron, John Whitmarsh and directors of NSF and NIH for organizing this meeting, and taxpayers for paying for it.

This talk will give an overview, big picture, with 3 concrete examples.

Mathematics Is Biology's Next Microscope, Only Better



- Microscope (late 17th C.) caused revolution in biology by revealing otherwise invisible, & previously unsuspected, worlds.
- Mathematics (broadly interpreted) can reveal otherwise invisible worlds in all kinds of data, not only optical.

Microscope

<http://www.healthmedialab.com/art/exhibits/leeuwenhoek.jpg>

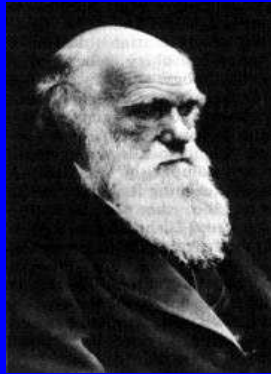
Mütter Museum, The College of Physicians of Philadelphia. Used with permission.

Specimen from the Visible Human Male - Head subset

http://www.nlm.nih.gov/research/visible/image/head_fresh.jpg

National Library of Medicine

“Mathematics seems to endow one
with something like a new sense.”
Charles Darwin



Bell, Eric Temple 1937 Men of Mathematics. New York: Simon and Schuster. Darwin quoted on p. 16.

Darwin

<http://www.probertencyclopaedia.com/j/Charles%20Darwin.jpg>

Permission granted by the web site.

Biology Is Mathematics' Next Physics, Only Better



- Physics stimulated enormous advances in mathematics, e.g. geometry, calculus.
- Biology can stimulate creation of new realms of mathematics.
- Is living nature qualitatively more heterogeneous than non-living nature?

Newton

<http://www.probertencyclopaedia.com/j/Sir%20Isaac%20Newton.jpg>

Mendel

<http://www.probertencyclopaedia.com/j/Gregor%20Mendel.jpg>

Permissions granted by the web site.

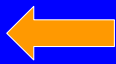
Originally, I asserted that the reason biology can stimulate the creation of new realms of mathematics is that life is more heterogeneous than non-living nature. But then I asked myself: How could I test quantitatively my intuitive feeling that living nature is qualitatively more heterogeneous than non-living nature? For example, there are 2,000-5,000 species of rocks and minerals in the Earth's crust, whereas there are perhaps 1,000 times more living species (somewhere between 3 million and 30 million – guesswork). But perhaps mineral species and biological species are not comparable. Is living nature more heterogeneous?

Shipman, J.T., Wilson, J.D. and Todd, A.W., An Introduction to Physical Science, 10th Ed., Houghton-Mifflin, 2003. Ch. 21 estimates 2000 minerals in Earth's crust.

Outline

- Past
 - biology
 - mathematics
- Present
 - landscapes of biology & applied math
 - examples
- Future
 - potential problems
 - opportunities

Outline

- Past
 - biology 
 - mathematics
- Present
 - landscapes of biology & applied math
 - examples
- Future
 - potential problems
 - opportunities

Biology since 1500

- External factor: Columbian exchange
(New World species, foods, diseases)
- Technical progress
 - quantitation (Harvey 1615 [1628])
 - microscope (Leeuwenhoek 1660-1700)
 - chemistry (Liebig 1855)
- Conceptual progress
 - anatomical observation (Vesalius 1543)
 - cell theory (Schleiden Schwann Virchow Weissman 1838-80)
 - evolution (Darwin Wallace 1859)
 - genes (Mendel Correns von Seysenegg deVries 1865-1900)

I summarize 500 years of biology in a single slide. The period since 1500 began with a major external shock: the discovery of New World species not previously described, the revolution in Old World agriculture from the introduction of New World cultivars, and the probable exchanges of smallpox and syphilis between the Old World and the New.

Important advances were technical and conceptual.

Technical: The use of the microscope by van Leeuwenhoek and many others to examine microbes revealed further new realms of life.

Conceptual: Vesalius' introduction of direct observation in anatomy challenged 14 centuries of dominance by Galen.

The most important conceptual advance was the Darwin-Wallace theory of evolution because it made biological variability a centerpiece of reality, not a shadow of some Platonic ideals. Darwin and Wallace finally shattered the grip of Platonism on Western thought.

Cell theory background:

<http://www.uwinnipeg.ca/~simmons/cellthry.htm>

1665 Robert Hooke - reported to the Royal Society of London his observation of "Cells" in a piece of cork. 1838-39 the general statement of "cell theory" receive general acceptance.

M.J. Schleiden and Theodor Schwann both published separate conclusions that expressed the principle of cell theory. "all living things are composed of living cells"

In 1855 Rudolf Virchow proposed an important extension of cell theory that "All living cells arise from pre-existing cells". ("Omnis cellula e cellula") This statement has become what is known as the "Biogenic law".

August Weissman 1880 "that cells living today can trace their ancestry back to ancient times"

Genes and chromosomes

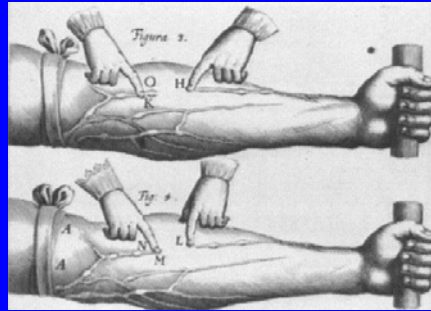
<http://webpages.marshall.edu/~adkinsda/B111OutlinesChromInhIntro.html>

Three botanists (Correns, von Seysenegg and deVries) independently discovered Mendel's principles of segregation and independent assortment.

William Harvey (1578-1659)



Exercitatio Anatomica
De Motu Cordis et
Sanguinis In
Animalibus (1628)



William Harvey, father of quantitative biology, used off-the-shelf mathematics (elementary arithmetic) to make a stunning discovery: blood circulates in the human body.

Studied at Padua 1600-02 while Galileo was active there. Knew blood circulated by 1615, but did not publish until 1628, partly out of concern for the unfavorable effect of such radical ideas on his practice.

Because Harvey was physician to the King, Harvey had access to the deer in the King's deer park. He also dissected executed criminals.

William Harvey

www.probertencyclopaedia.com/j/William%20Harvey.jpg

Permission granted by the web site.

Arms

phy025.lubb.ttuhsu.edu/Figures/circulation.gif

Image courtesy of the National Library of Medicine.

Harvey showed that blood circulates

- Galen (2d C.): blood ebbs & flows, pumped by arteries; heart is passive.
- Harvey (1615): heart & veins have 1-way valves; flow is unidirectional.
- Left ventricle (dead) holds 2 oz. ~ 60 ml.
- $>1/8$ - $1/4$ of blood is expelled per stroke.
- Heart beats 60-100 times/minute.
- $\therefore 60\text{ml} \times 1/8 \times 60 \text{ beats/min} \times 60 = 27 \text{ l/h.}$
- Average human has 5.5 l blood.

[Marcello Malpighi (1628-94) saw capillaries.]

Prior theory (for 14 centuries!) was that blood was created from food, that arteries were the active agent of pumping, that blood ebbed and flowed, like the tides, through blood vessels, and that the heart was a passive receptacle.

Harvey proved that the hourly output of the heart was many times more than the total blood volume of the body. All that output couldn't come from ingested food, because people don't eat 27 l/h of food. So the blood that goes out must come back again to the heart, i.e., blood circulates. Harvey could not see the smallest vessels (capillaries) through which blood returns from the arteries to the veins, but his quantitative argument proved they must exist.

By end of 17th century, microscopists observed capillaries directly. Van Leeuwenhoek confirmed Malpighi's observation of capillaries.

Taylor, F. Sherwood 1949 A Short History of Science and Scientific Thought. New York: Norton. Harvey's reasoning is quoted on p. 103.

Harvey took account of uncertainty of observations. E.g., he said that ventricle holds 2 oz., with a RANGE from 1.5 to 3 oz. Expulsion factor RANGED from $1/8$ to $1/4$. He was conscious of uncertainty.

This must be one of the most beautiful examples of the tremendous power of simple, off-the-shelf mathematics to illuminate a biological problem. Harvey predicted the existence of capillaries (to return the blood from the arteries to the veins) from pure theory, more than half a century before they were observed.

Outline

- Past
 - biology
 - **mathematics** 
- Present
 - landscapes of biology & applied math
 - examples
- Future
 - potential problems
 - opportunities

Mathematics since 1500

- Geometry & topology
 - analytic geometry (Descartes 1637)
 - non-Euclidean geometries (1823-1830)
- Algebra
 - roots of equations, group theory, symmetry
 - linear algebra (19th-20th C.)
- Analysis: *modern mathematical thought*
 - probability (Pascal Fermat 1654 DeMoivre Laplace Gauss 1733-1809 normal curve)
 - calculus: theory of limits (Newton 1666 Leibniz)
- External factor: computers (war, business)

Descartes married the Greek and the Arabic traditions of mathematical thought by putting a coordinate system on the featureless geometric plane of the Greeks and showing that algebraic equations have simple geometric interpretations, and simple geometric figures have simple algebraic representations. Any biologist who has ever plotted (x, y) data is using the invention of Descartes.

Modern thought is separated from classical thought by the calculus (the theory of limits and continuous functions) and the theory of (continuous) probability. Neither theory has any counterpart in classical thought. No person who has not understood at least the rudiments of both of these theories has entered modern thought.

<http://www.stat.wvu.edu/SRS/Modules/Normal/normal.html>

"History

The normal curve was developed mathematically in 1733 by DeMoivre as an approximation to the binomial distribution. His paper was not discovered until 1924 by Karl Pearson. Laplace used the normal curve in 1783 to describe the distribution of errors. Subsequently, Gauss used the normal curve to analyze astronomical data in 1809. The normal curve is often called the Gaussian distribution. The term bell-shaped curve is often used in everyday usage."

Whereas biology received its main external shock (discovery of New World) around 1500, math received its main external shock (computers) toward the end of the most recent 5 centuries, driven by the demands of war and business. Hollerith punched cards originated in the control of looms for weaving. The demand for ordnance tables and bomb calculations drove the early development of computers.

Mathematics arising from biological problems (1)

- Age structure of stable populations (Euler 1760)
- Logistic equation for limited population growth (Verhulst 1838)
- Branching processes, extinction of family names (Galton 1889)
- Correlation (K. Pearson 1903)

This list is not to identify all the many places where mathematics has been useful in biology, but to give examples where a biological challenge or problem led to the creation of new mathematics. To the question, what can biology do for mathematics?, here are some answers. The list is intended to be illustrative, not exhaustive, and is not intended as a scholarly history. For example there is a history of game theory before von Neumann (1944), and Pearson's development of the correlation coefficient rested on earlier work by Galton. Many of these developments blend smoothly into their antecedents and later elaborations.

Keyfitz, *Introduction to the Mathematics of Population* gives references to Euler, Verhulst.

K. Pearson and A. Lee, 1903 On the laws of inheritance in man. *Biometrika* ii, 357-462.

Mathematics arising from biological problems (2)

- Markov chains, statistics of language (Markov 1906)
- Hardy-Weinberg equilibrium (1908)
- Analysis of variance, design of agricultural experiments (Fisher 1920s)
- Dynamics of interacting species (Lotka 1922 Volterra 1926-37)

Mathematics arising from biological problems (3)

- Birth process (Yule 1925), birth and death process (D.G. Kendall 1948)
- Traveling waves in genetics (Fisher; Kolmogorov Petrovsky Piscounov 1937)
- Game theory (von Neumann 1944)
- Distribution for estimating bacterial mutation rates (Luria Delbruck 1943)
- Morphogenesis (Turing 1952)

Noah Rosenberg <noahr@hto-b.usc.edu> Yule's birth process (Phil Trans R Soc Lond Ser B [1925], 21-87): develops the pure birth process to study the distribution of the number of species in a genus.

M. S. Bartlett *Stochastic Processes* gives ref. to DG Kendall 1948

R. A. FISHER, The wave of advance of advantageous genes, *Annals of Eugenics* **7**, 353-369 (1937).

A. KOLMOGOROV, I. PETROVSKY, AND N. PISCOUNOV, Etude de l'équation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique, *Moscow University Bull. Math.* **1**, 1-25 (1937).

J. A. MURRAY, *Mathematical Biology* (Springer-Verlag, Berlin, 1998).

M. BRAMSON, Convergence of solutions of the Kolmogorov equations to traveling waves, *Memoirs of the American Mathematical Society* **44** (1983).

Example of recent work on this equation:

<http://www.math.rochester.edu/people/faculty/cmlr/Preprints/noisy-fkpp.pdf>

Noah Rosenberg <noahr@hto-b.usc.edu> Luria-Delbruck distribution (Genetics 28, 491-511 [1943]): initially a method for estimating bacterial mutation rates, the distribution has since been studied as an entity in itself (Kepler and Oprea, Theor Pop Biol 59, 41-48 [2001]).

Mathematics arising from biological problems (4)

- Diffusion equation for gene frequencies (Kimura 1954)
- Circular interval graphs, genetic fine structure (Benzer 1959)
- Threshold functions of random graphs, models of communication networks or “even of organic structures of living matter” (Erdős-Rényi 1960)

Noah Rosenberg <noahr@hto-b.usc.edu> “Kimura's solution of the diffusion equation for gene frequencies (papers 1-4 in Kimura "Population Genetics, Molecular Evolution, and the Neutral Theory" [1994]): ... some of the mathematics was independently derived by Kimura, leading to further interest in diffusion methods ...”

Crow and Kimura, *An Introduction to Population Genetics Theory*, gives detailed bibliography.

<http://dimacs.rutgers.edu/dci/2001/abstractswk1right.html>

Fred S. Roberts Rutgers University From Genes to Archaeological Digs and from Traffic Lights to Childhood Development: The Many Applications of Interval Graphs

“The concept of interval graph was introduced by the Hungarian mathematician Hajos in connection with a scheduling problem and independently by the geneticist Seymour Benzer in connection with the problem of understanding the makeup of the fine structure inside the gene. Since then, this one simple idea has had applications in archaeology, developmental psychology, utility theory in economics, traffic light phasing, ecology, and many other areas, and has given rise to some fascinating mathematical theories and algorithms.”

Mathematics arising from biological problems (5)

- Sampling formula for haplotype frequencies (Ewens 1972)
- Coalescent, genealogy of populations (Kingman 1982)

Noah Rosenberg <noahr@hto-b.usc.edu>

“Ewens sampling formula (*Theoretical Population Biology* 3, 87-112 [1972]): a probability distribution that arises from considering haplotype frequencies in a population. Apparently the distribution has since appeared in connection with various other combinatorial problems (Johnson, Kotz, and Balakrishnan "Discrete Multivariate Distributions" 1997, p. 232-246).”

<http://www.stats.ox.ac.uk/mathgen/coalescent.html>

“Coalescent methods

One of the most important recent developments in population genetics modelling is the use of coalescent or genealogical methods. These methods focus on modelling the genealogy (ancestral history) of a random sample of genes from the population, using a stochastic process known as the coalescent (Kingman, 1982). An understanding of the genealogy can give both a qualitative and quantitative understanding of the patterns of variation in genetic data, and use of coalescent methods leads to efficient simulation methods, and to inference techniques which make full use of all information in the data.

Selected references Donnelly, P. and S. Tavaré (1995). Coalescents and genealogical structure under neutrality. *Ann. Rev. Genet.*, 29:401-421 Hudson, R.R. (1991). Gene genealogies and the coalescent process. In *Oxford Surveys of Evolutionary Biology*, ed. D. Futuyma and J. Antonovics, 7:1-44. Oxford: Oxford University Press.

Kingman, J.F.C. (1982). On the genealogy of large populations. *J. Appl. Prob.*, 19A:27-43 Kingman, J.F.C. (1982). The coalescent. *Stochastic Processes and their Applications*, 13:235-248”

Outline

- Past
 - biology
 - mathematics
- Present
 - landscapes of biology & applied math 
 - examples
- Future
 - potential problems
 - opportunities

I offer a map of the landscapes of biology and applied mathematics to provide a framework for classifying and relating the many diverse efforts that use mathematics in biology.

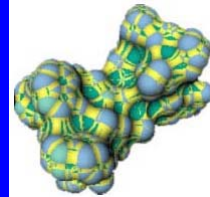
The biological landscape

Questions

- Structure(s)
 - How is it built?
- Function(s)
 - How does it work?
- Pathology(ies)
 - What goes wrong?
- Repair(s)
 - How is it fixed?
- Origin(s)
 - How did it begin?

Domains

- Molecules
- Cells
- Tissues
- Organs
- Individuals
- Populations
- Communities, ecosystems
- Biosphere



Gramicidin molecule

<http://www.ticam.utexas.edu/CCV/projects/Angstrom/gramicidin.jpg>

Permission granted by Chandrajit Bajaj. "Dynamic Maintenance and Visualization of Molecular Surfaces", C. Bajaj, V. Pascucci, R. Holt, A. Netravali, Fourth Issue of Discrete Applied Mathematics on Computational Molecular Biology", 2002.

<http://www.ticam.utexas.edu/CCV/projects/Angstrom/>

Mouse <http://ink.primare.wisc.edu/~assay/mouse.jpg>

Permission granted by Fritz Wegner, Wisconsin National Primate Research Center

Corals <http://www.rockefeller.edu/pubinfo/012103a.jpg>

Permission granted by LeRoy Headlee

Mountain landscape photograph by Joel E. Cohen from High Point, Ulster County, New York 2001. Commercial use without permission is prohibited.

The questions correspond to the rows of a matrix, while the domains identify the columns of the matrix. Many biological research problems can be classified by the combination of question(s) and domain(s). Many studies involve more than one of these questions and more than one of these domains.

The applied mathematical landscape

- Data structures
- Algorithms
 - analyze data
 - analyze models
- Theories & models, including all pure math
 - analyze data
 - analyze ideas
- Computers & software
 - embody mathematical knowledge
 - interface with humans (vision, speech)
 - compute

These four main headings of the applied mathematical landscape might be thought of as lying at the vertices of a tetrahedron, each strongly interacting with all the others.

Algorithms that analyze data: least squares

Algorithms that analyze models: finding eigenvectors of a matrix

From: Zaida Luthey-Schulten <schulten@scs.uiuc.edu>

“Mathematics Unlimited - 2001 and Beyond” Engquist and Schmid, Editors (Springer Press). Top Ten Algorithms that had the most influence on problems in science and engineering in the previous century. The original list appeared in the first issue of the new millennium in “Computing in Science and Engineering Journal”: Top Ten Algorithms

1. Metropolis Algorithm (Monte Carlo)
2. Simplex Method for Linear Programming
3. Krylov Subspace Methods - to solve sparse matrices in linear systems of equations
4. QR algorithms to decompose matrix - eigenvalue problems
5. Optimization Fortran Compiler
6. Decomposition Approaches to Matrix Computations - LU, Cholesky, Schur,...
7. Quicksort
8. Fast Fourier Transform
9. Fast Multipole methods - For problems in electrostatics
10. Integer Relation Detection Methods - For classification of Feynman diagrams in quantum field theory

Several authors including I think that the list should also include Molecular Dynamics, Dynamic Programming Algorithms, Hidden Markov Methods, Genetic Algorithms”

The landscape of biology and mathematics

The landscape of research in mathematics and biology contains all combinations of a problem from the matrix of biological problems and problem areas from applied mathematics.

The landscape of research in the intersection of mathematics and biology can be thought of as the Cartesian product of the matrix of problems from biology times the tetrahedron of problems from applied mathematics: each topic of research combines a question, a biological domain, and some set of mathematical techniques or results.


Example 1 comes from the domain:

How does it work x cells x algorithms & computer software

Example 2 comes from the domain:

How is it built x ecological community x data structure

Outline

- Past
 - biology
 - mathematics
- Present
 - landscapes of biology & applied math
 - examples: complete data are published 
- Future
 - potential problems
 - opportunities

Example 1: integrate gene expression & molecular pharmacology

- Scherf U, 16 others. A gene expression database for the molecular pharmacology of cancer. *Nature Genetics* 2000 Mar; 24(3):236-44.

Scherf U, Ross DT, Waltham M, Smith LH, Lee JK, Tanabe L, Kohn KW, Reinhold WC, Myers TG, Andrews DT, Scudiero DA, Eisen MB, Sausville EA, Pommier Y, Botstein D, Brown PO, Weinstein JN. A gene expression database for the molecular pharmacology of cancer. *Nat Genet.* 2000 Mar;24(3):236-44.

<http://discover.nci.nih.gov/arraytools/>

“We used cDNA microarrays to assess gene expression profiles in 60 human cancer cell lines of the National Cancer Institute’s drug discovery program. The accompanying paper (Ross et al.) describes the biological implications; here, using the same microarray data, we link the bioinformatics with chemoinformatics by correlating gene expression and drug activity patterns in the 60 cell types. Clustering the cells on the basis of gene expression yields a picture very different from that obtained when the cells are clustered on the basis of their response to drugs. Gene-drug relationships for the important clinical agents 5-fluorouracil and L-asparaginase exemplify how variations in the transcript levels of particular genes can relate to mechanisms of drug sensitivity and resistance. This is the first study to integrate large databases on gene expression and molecular pharmacology.”

Results

- “used cDNA microarrays to assess gene expression profiles in 60 human cancer cell lines of the National Cancer Institute’s drug discovery program.”
- “link the bioinformatics with chemoinformatics by correlating gene expression and drug activity patterns in the 60 cell types. Clustering the cells on the basis of gene expression yields a picture very different from that obtained when the cells are clustered on the basis of their response to drugs.”

Data <http://discover.nci.nih.gov/arraytools/>

Drug activity matrix A

1400 drugs x 60 cell lines from human cancers of various organs (including 118 drugs with “known mechanism of action”)

a_{dc} = activity of drug d in suppressing growth of cell line c

a_{dc} = sensitivity of cell line c to drug d

Gene expression matrix T (“T” for target)

1375 genes x same 60 cell lines

t_{gc} = relative abundance of mRNA transcript of gene g in cell line c

t_{gc} = cell line c’s expression of gene g

Drug activity matrix A = (a_{dc})

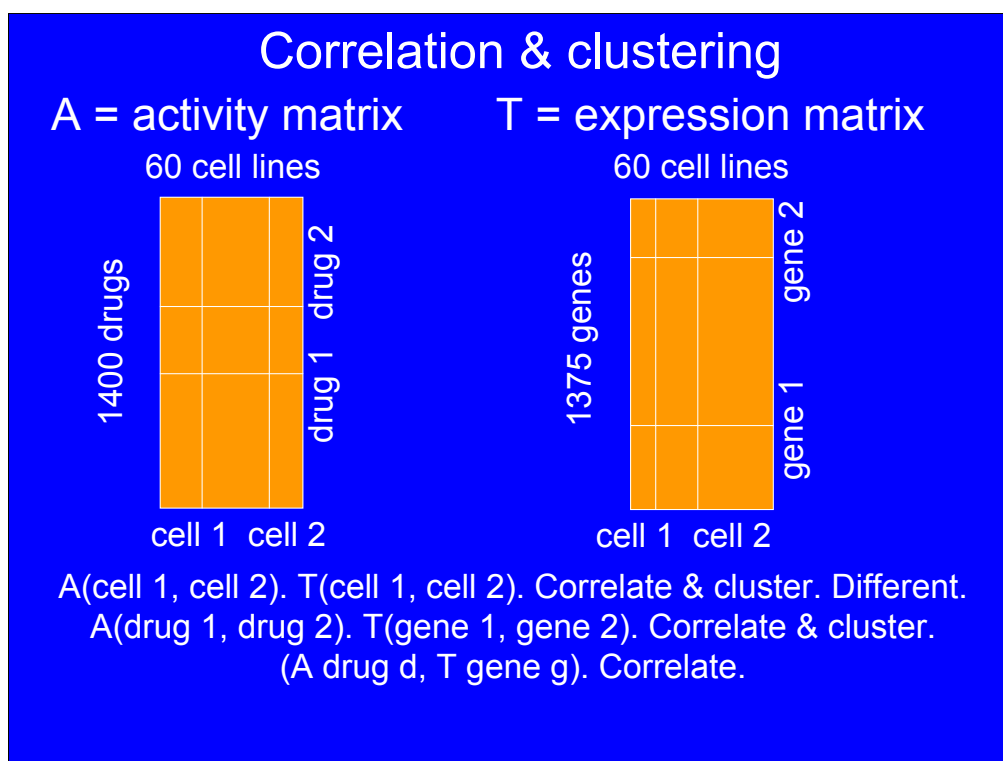
a_{dc} = $-\log_{10}$ concentration of drug d required to inhibit growth of cell line c by 50% compared to controls

1400 drugs x 60 cell lines from human cancers of various organs

Gene expression matrix T = (t_{gc})

t_{gc} = $+\log_2$ relative abundance of mRNA transcript of gene g in cell line c

1375 genes x same 60 cell lines

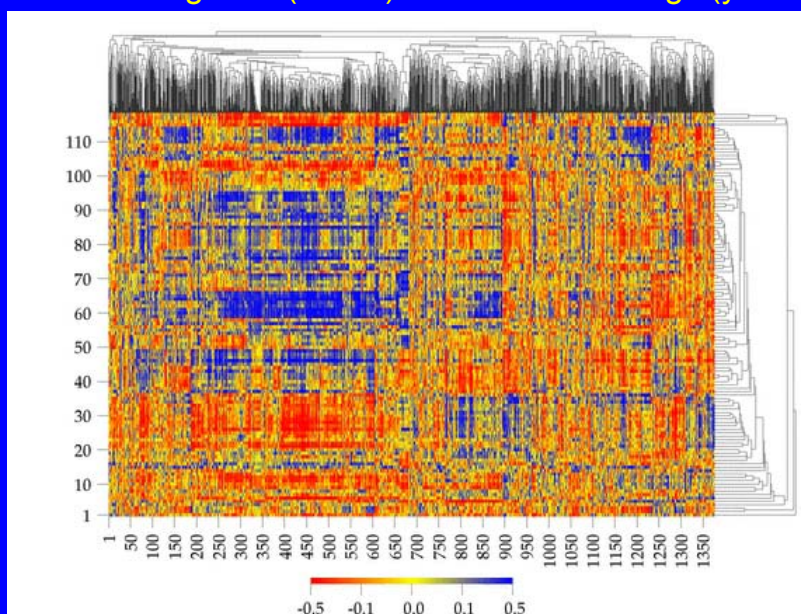


Given two matrices that share the same set of column headings, there are precisely five possible sets of correlations that can be calculated. This paper calculated all five.

1. The correlation between different columns of the activity matrix leads to a clustering of cell lines according to their responsiveness to different drugs.
2. The correlation between different columns of the expression matrix leads to a clustering of cell lines according to their gene expression. The latter clustering closely corresponds to the organ from which the tumor came (leukemia, lung, CNS, etc.), and differs substantially from the clustering of cell lines by drug activity.
3. The correlation between different rows of the activity matrix leads to a clustering of drugs according to their activity patterns across all cell lines.
4. The correlation between different rows of the expression matrix leads to a clustering of genes according to the pattern of amount of mRNA expressed.
5. Finally, the correlation between a row of the activity matrix and a row of the expression matrix describes the positive or negative covariation of drug activity with gene expression. Scherf et al. plotted the activity-expression correlations for each combination of drug and gene after clustering the drugs and genes in steps 3 and 4.

Clustered image map

gene expression-drug activity correlations plotted as a function of clustered genes (x-axis) and clustered drugs (y-axis)



Source of CIM (clustered image map)

http://discover.nci.nih.gov/external/CIM_example3/cgi_user_matrix.html

Used by permission of John N. Weinstein.

Updated Thu Apr 27 07:42:34 EDT 2000. Clustered image plot of X = 1376 bigdata genes, and Y = 118 best median averaged drugs, correlated on 60 cells. This image is more recent than the image in the published paper.

Use of this: if gene expression is high, and correlation with drug activity is high, then gene can be used as a marker for tumor cells likely to be treated effectively by the drug. If correlation with drug activity is negative, then marker gene may indicate when use of drug is contradicted.

The rectangular blocks of color identify functionally related genes and functionally similar drugs.

Questions:

Does one need control non-tumorous cell lines?

Does lack of independence of correlation measures affect statistical significance of correlations? Does bootstrap solve all problems?

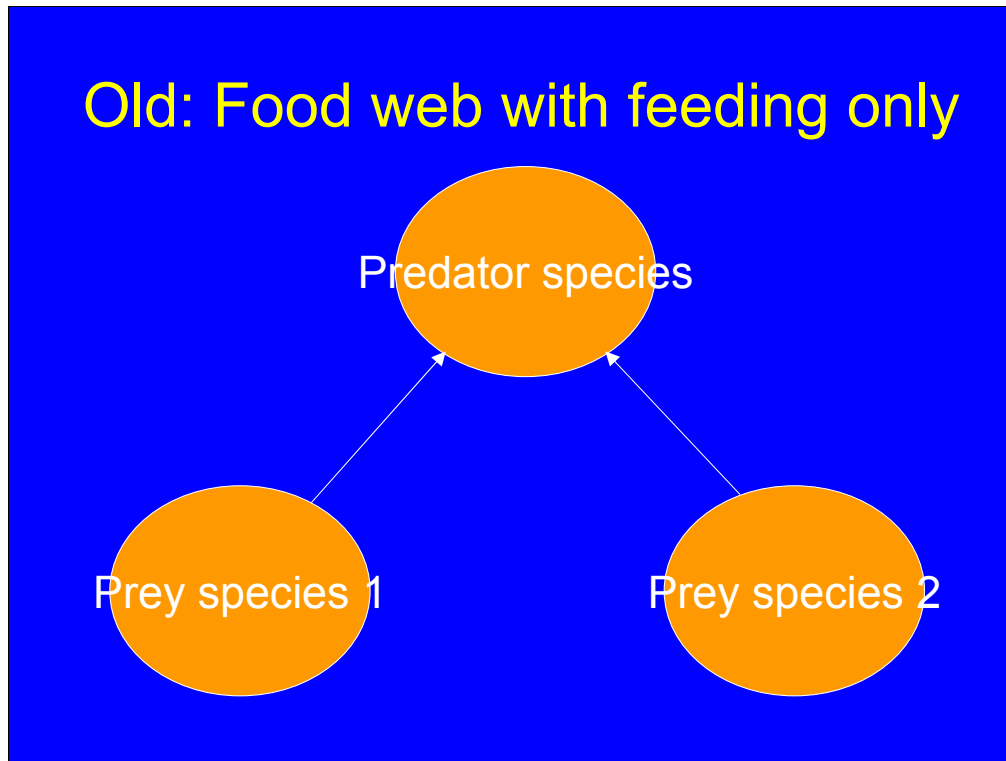
Do in vitro drug activity (inhibition of protein synthesis), and in vitro gene expression of non-drug treated cell lines, predict in vivo response of tumors to drugs?

How can measurements from cDNA microarrays be extended to combined measurements of DNA, RNA and proteins?

Larger picture: how to embed people in genealogies & changing environments?

Example 2: integrate
food webs (attribute of ecological
communities) with
body size (attribute of individuals)
& abundance (attribute of
populations)

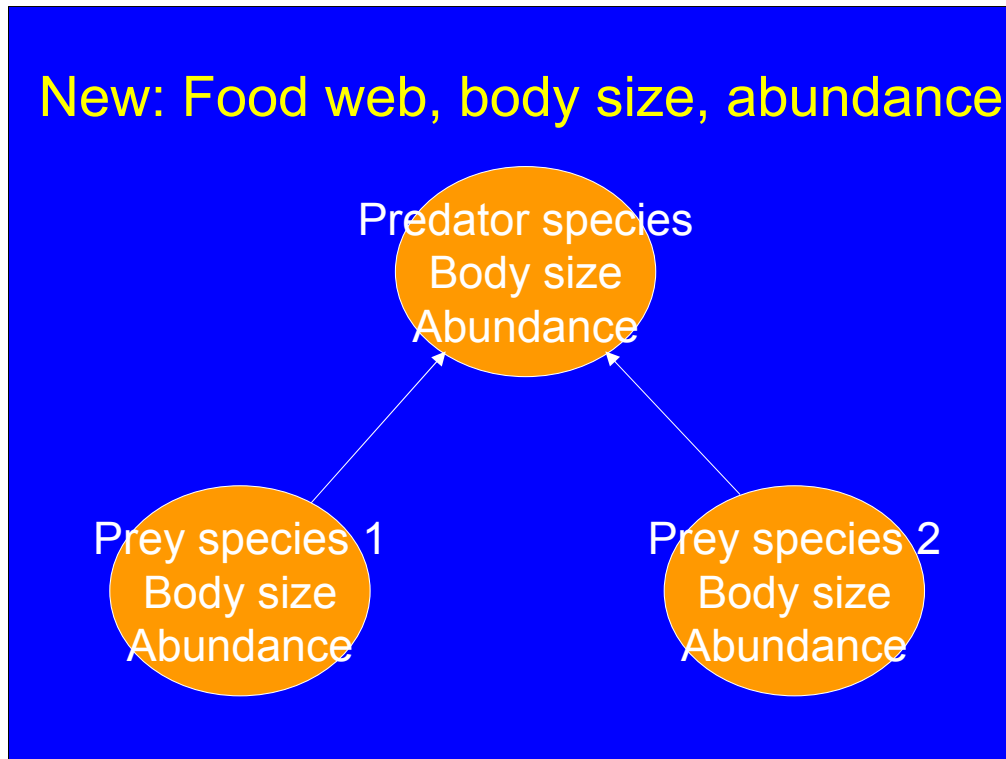
Joel E. Cohen, Tomas Jonsson, Stephen R. Carpenter,
Ecological community description using the food web,
species abundance, and body size. *Proc. National
Acad. Sci. USA* 100(4):1781-1786, 18 February 2003



Arrow shows direction of flow of food FROM prey TO predator.

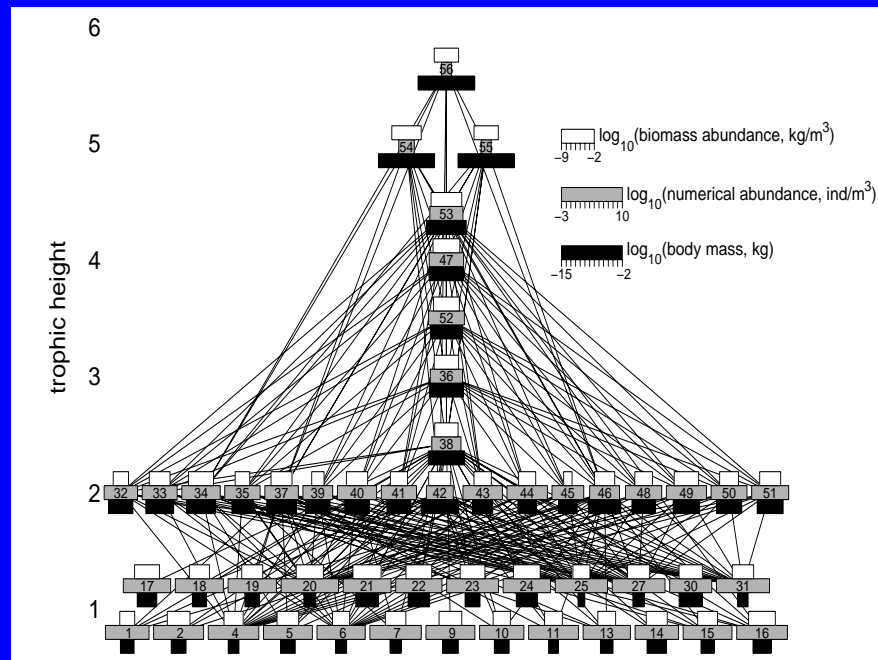
Camerano, Lorenzo 1880 Dell'equilibrio dei viventi merce la reciproca distruzione. Atti della Reale Accademia delle Scienze di Torino 15:393-414. (Translation by Claudia M. Jacobi, edited by J. E. Cohen) On the equilibrium of living beings due to reciprocal destruction. In: *Frontiers of Theoretical Biology. Lecture Notes in Biomathematics* 100, ed. Simon A. Levin, pp. 360-380. New York: Springer-Verlag 1994

This first, and still widespread, graphical representation of a food web derives from the only picture in Darwin's *Origin of Species*, which is a tree for the evolution of life forms.



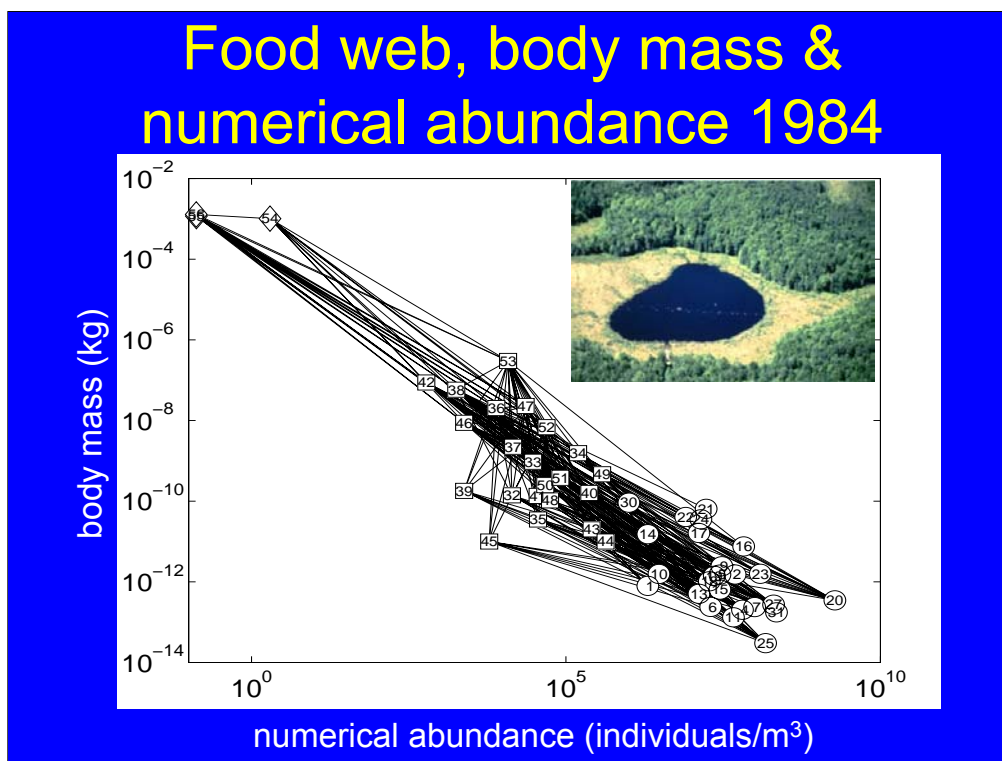
We added 2 quantitative facts about each species. While the food web is an attribute of an ecological community, the average body size characterizes individuals (weighted by the distribution in a population) and the abundance characterizes populations. More generally, one could add any number of quantitative or qualitative facts about each species (e.g., chemical composition) as well as quantitative or qualitative facts about each link (e.g., amounts of energy or specific elements flowing from one species to another).

Tuesday Lake 1984



Copyright 2003 National Academy of Sciences, U.S.A. Reproduced here on authority of Joel E. Cohen, co-author.

The unlumped food webs of Tuesday Lake in 1984. The width of the black, grey and white horizontal bars shows, respectively, the \log_{10} body mass (kg), \log_{10} numerical abundance (ind./m³) and \log_{10} biomass abundance (kg/m³) of each species. The vertical positions of the species show trophic height. Basal species have a trophic height of unity by definition, but to allow for wider, non-overlapping bars the vertical positions of the basal species have been adjusted around unity. The horizontal position is arbitrary. Isolated species are omitted. Species with a trophic height of 1 are phytoplankton, 3 species with a trophic height of 5 or more are fish, species with intermediate trophic heights are zooplankton.



Copyright 2003 National Academy of Sciences, U.S.A. for black & white graphic.
Reproduced here on authority of Joel E. Cohen, co-author.

Photograph of Tuesday Lake 0.9 ha surface area

limnology.wisc.edu/slideshow/water/lake4.jpg

Image courtesy of S. Carpenter, Center for Limnology, University of Wisconsin-Madison.

The Tuesday Lake food web in 1984, in the numerical abundance-body size space. The center of a node gives the numerical abundance (individuals/m³) and average body mass (kg) for the species identified by the number within the node. Edges connect species that interact trophically. Isolated species are omitted. Circles = phytoplankton, squares = zooplankton, diamonds = fish.

Body mass times numerical abundance gives biomass. Line of slope -1 is locus of constant biomass. Such a line fits right through the data.

10 orders of magnitude variation in N

12 orders of magnitude variation in M

5 orders of magnitude variation in $B = MN$.

When food web is used to adjust species B by size of available prey B, fit to straight line improves. Taking account of food web improves orderliness of the data.

Outline

- Past
 - biology
 - mathematics
- Present
 - landscapes of biology & applied math
 - examples
- Future
 - potential problems 
 - opportunities

First the bad news, then the good news: problems, then opportunities.

Potential problems

1. Educating scientists
2. Intellectual property
3. National security



Potential problems (1) educating scientists

- K-16 education in math, science
 - U.S. math & science education falls behind that of other industrial nations.
- graduate, postdoc cross-training
- diversity: gender, ethnicity, nationality (visas)
- educating peer reviewers to approve exploratory research of high quality

“U.S. Urged to Revise Visa Restrictions for Foreign Scientists

The federal government must revise current visa restrictions for foreign scientists, engineers and qualified students if the United States is to maintain the vitality and quality of its research, said the presidents of the National Academies in a recent statement. "Efforts by our government to constrain the flow of international visitors in the name of national security are having serious unintended consequences for American science, engineering and medicine," they said. "We ask the Department of State and its consular officials to recognize that, in addition to their paramount responsibility to deny visas to potential terrorists, the long-term security of the United States depends on admitting scholars who benefit our nation." “

Statement:

<http://www4.nationalacademies.org/news.nsf/isbn/s12132002?OpenDocument>

School children

<http://gimp-savvy.com/cgi-bin/img.cgi?ufwsaX11fXH8dr6154>

Photo Credit: US Fish and Wildlife Service

Potential problems (2) intellectual property

- Science as a potlatch culture
- Bayh-Dole 1980
 - Rai & Eisenberg, *Amer. Scientist* 91:52, 2003
- Tragedy of anti-commons: *Madey v. Duke*
 - Duke Petition for Writ of Certiorari to U.S. Supreme Court: “The possibility that the patent system could stifle or even stymie the progress of biotechnology and other important fields of research is both real and profound.”



Supreme Court

http://a255.g.akamaitech.net/7/255/2422/08jun20011500/www.supremecourt.us.gov/images/court_front_med.jpg

Image by Franz Jantzen, Collection of the Supreme Court of the United States. Used with permission.

In scientific culture, the bigger the gift of ideas or discoveries that a scientist makes to the common pool, the higher that scientist's prestige as a scientist. This culture conflicts with the proprietary impulse behind protection of intellectual property.

Bayh-Dole legislation was intended to enhance the commercial exploitation of university-generated ideas, by giving universities and scholars at universities the right to patent discoveries. But the scrambling for IP rights it has induced may inhibit the research that generates ideas.

Administrators of grants of public monies should assure that public interests in the results are protected.

Potential problems (3) national security

- Win-win domains
 - EPA biowarfare monitoring
 - foot & mouth disease
 - smallpox inoculation strategy
- Good (openness) vs. good (defense)
- Good (privacy) vs. good (security): databases, biomarkers (SNPs)
- “Sensitive but unclassified” information
 - who does the research? (non-US?)
 - with what publication rights/obligations?



Scientists and national defenders can collaborate by supporting and doing open research on optimal design of monitoring networks and mitigation strategies for all kinds of biological attacks, including anthrax; improving prevention, early detection and response to foot and mouth disease and other hazards to the food supply; and optimal design of smallpox inoculation strategies. E.g., Lawrence M. Wein, David L. Craft, and Edward H. Kaplan 2003 Emergency response to an anthrax attack. Proc. National Acad. Sci. 100 (7): 4346–4351 April 1, 2003 www.pnas.org/cgi/doi/10.1073/pnas.0636861100
 Commentary in same issue: G.F.Webb 2003 A silent bomb: The risk of anthrax as a weapon of mass destruction

But sometimes values conflict: openness (of scientific methods or biological reagents, for example) may pose security risks in the hands of terrorists. Personal privacy (with respect to genetic markers, biometric characteristics) may impede efforts to track security suspects. How should researchers and public officials trade off these conflicting goods?

Problem of conserving privacy when disparate databases are connected: physician payment and disease diagnosis, or health and law enforcement

Inadequate or improper definition of “sensitive but unclassified” information may have a serious chilling effect on many varieties of mathematical and biological research.

Army image in the public domain

<http://www.legionxxiv.org/images/photo02.jpg>

Outline

- Past
 - biology
 - mathematics
- Present
 - landscapes of biology & applied math
 - examples
- Future
 - potential problems
 - opportunities 

Mathematics can help biologists grasp problems that are otherwise

- Too big
 - biosphere
- Too slow
 - macro evolution
- Too remote in time
 - early extinctions
- Too complex
 - brain
- Too small
 - molecular structure
- Too fast
 - photosynthesis
- Too remote in space
 - life at extremes
- Too dangerous or unethical
 - epidemiology of infectious agents

Sigma Xi, NSF's National Computational Science Institute, and Shodor Education Foundation, Inc. of Durham North Carolina are running computer model workshops to help science students investigate in the classroom topics that are otherwise invisible because they are

too small: JEC ion pores in the cell membrane

too large: JEC oceanic food webs

too fast: photosynthesis

too slow: JEC macro evolution of major clades; large-scale ecosystem responses to climatic change

too complex: JEC brain

too dangerous: JEC alternative vaccination strategies in response to intentional smallpox introductions; Chagas disease model for removal of dogs from households without spraying insecticide to remove vectors (Cohen, J. E. and Gürtler, Ricardo E. 2001 Modeling household transmission of American trypanosomiasis. *Science* 293(5530):694-698; Web supplement at

<http://www.sciencemag.org/cgi/content/full/293/5530/694/DC1> 27 July);

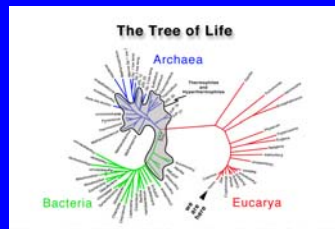
invasions of biological species

<http://www.computationalscience.net/press/pdf/relationship.pdf> accessed 14 Feb 2003

http://www.ncsc.org/news/pr/Shodor_final.pdf accessed 14 Feb 2003

Biological challenges (1)

- Understand cells, their diversity within & between organisms, & their interactions with biotic & abiotic environments
- Understand brain, behavior & emotion
 - Why do or don't people have children?
- Replace tree of life by network to represent lateral transfers of genes & genomes



Tree of life

<http://www.astro.washington.edu/endsofworld/tree%20of%20life%20%20final.jpg>

Permission granted by Don Brownlee

For rare examples of phylogenies that recognize horizontal transfer of genes and acquisition of genomes, see

Delwiche, C.F. 1999. Tracing the web of plastid diversity through the tapestry of life. *American Naturalist* 154:S164-S177. Revised figure <http://www.life.umd.edu/labs/Delwiche/pubs/endosymbiosis.gif>

See p. 878, Fig. 4, of Delwiche, C.F., and J. D. Palmer. 1996. Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. *Mol. Biol. Evol.* 13:873-882.

Margulis, Lynn and Sagan, Dorion 2002 *Acquiring Genomes: A Theory of the Origins of Species*. Basic Books, New York.

Delwiche, C.F. 2000. Griffins and chimeras: evolution and horizontal gene transfer. *Bioscience* 50:85-87 (book review of *Horizontal Gene Transfer*, 1998, M. Syvanen and C.I. Kado, eds., Chapman & Hall, London).

Delwiche, C.F. 2000. Gene Transfer Between Organisms. Pp. 193-197 in: *McGraw-Hill 2001 Yearbook of Science & Technology*. McGraw-Hill, New York.

Biological challenges (2)

- Couple atmospheric, terrestrial & aquatic biosphere with global physicochemical processes



- Monitor living systems to detect large deviations
 - epidemics natural or induced
 - physiological or ecological pathologies

Tornado <http://www.spc.noaa.gov/faq/tornado/alfalfa.jpg>

Photo courtesy of NSSL

Mathematical challenges (1)

- Understand computation: gaining insight & proving theorems from numerical computation & agent-based models
- Model multilevel systems, e.g., cells in people in human communities in physical, chemical, & biotic ecologies
- Understand uncertainty & risk by integration of frequentist, Bayesian, subjective & other theories of probability

The purpose of computing is insight, not numbers.

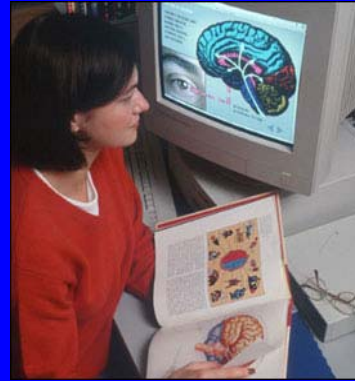
-- Richard W. Hamming, *Introduction to applied numerical analysis*, McGraw-Hill 1971, p31.

<http://www.cse.psu.edu/~schwartz/quotes.html>

Despite nearly 3 centuries of phenomenal progress, we are still at the very beginning of a true understanding of probability.

Mathematical challenges (2)

- Understand data mining, simultaneous inference (beyond Bonferroni)
- Set standards for clarity, performance, publication & permanence of software & computational results



Image

<http://online.rit.edu/images/woman-computer.jpg>

Photo credit: A. Sue Weisler, Rochester Institute of Technology. Used with permission.

Thank you.
Your thoughts?
cohen@rockefeller.edu



Photograph of Lava Lake, Montana, by Joel E. Cohen 2001. Commercial use without permission is prohibited.

Acknowledgments

I am grateful for helpful suggestions from Mary Clutter, Charles Delwiche, Bruce A. Fuchs, Yonatan Grad, Alan Hastings, Kevin Lauderdale, Zaida Luthey-Schulten, Daniel C. Reuman, Noah Rosenberg, Michael Pearson and Samuel Scheiner, support from U.S. National Science Foundation grant DEB 9981552, and the hospitality of Mr. and Mrs. William T. Golden during this work. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Additional useful links and references as of 26 February 2003:

<http://www.maa.org/mtc/> with link to:

John R. Jungck, "Ten equations that changed biology: mathematics in problem-solving biology curricula," *Bioscene* 23(1)11-36, May 1997

<http://www.nsf.gov/pubsys/ods/getpub.cfm?nsf02125>

http://www.nigms.nih.gov/about_nigms/cbcb.html

<http://www.bisti.nih.gov/>